(+91) 9971733565
Mumbai, Maharashtra, India
yadneshujwal@gmail.com

Yadnesh Salvi

Senior Data Scientist

GitHub: yadneshSalvi
LinkedIn: yadnesh-salvi-bb5151ba

AI/ML specialist with expertise in NLP, GenAI, LLMs, and RAG systems. Proven track record of implementing production-scale AI solutions that reduce costs and accelerate business processes. Passionate about translating AI research into practical applications.

Professional Experience

Senior Data Scientist - ContractPodAI - Mumbai **July, 2021 - Present**

- **Legal Deep Research AI Agent:**
- Built an **end-to-end AI agent** enabling legal teams to query hundreds of thousands of legal documents.
- Implemented a scalable **document ingestion pipeline** with intelligent chunking, vectorization, and metadata extraction.
- Developed the **core AI agent logic using Langgraph**, featuring query clarification, metadata-driven search space reduction, parallelized sub-query execution, and answer validation. Initiated front-end POC with **Copilotkit in React**.
- **Legal Contract Review with GenAI:**
- Significantly reduced contract review time **from days to minutes** for legal teams by developing AI powered solution.
- Architected a robust system design leveraging **FastAPI, Langchain, MongoDB, Qdrant**, and **Azure Functions**.
- Elevated redlining accuracy by over 60% using **advanced RAG approach** outperforming naive prompting methods.
- Created **custom evaluation datasets** using CUAD benchmark and Langsmith to quantitatively measure RAG pipeline accuracy.
- Developed comprehensive LLM interaction logging system resulting in faster troubleshooting of **LLM hallucinations**.
- Reduced production bugs by **enhancing the CI pipeline** with automated code quality checks and comprehensive pytest testing.
- **Information Extraction and Summarization:**
- Streamlined legal research by developing a scalable solution for rapid extraction and concise summarization.
- Integrated a **robust citation system** that displays exact source locations within legal documents, enhancing the verifiability.
- Optimized responsiveness by leveraging **asyncio** and **Langchain CallbackHandlers** to stream diverse data formats.
- Reduced document summarization latency by over 70% implementing **asynchronous map-reduce approach**.
- **Fine-tuned Llama-3 8B and GPT-4o** models on custom datasets and served at scale through vLLM on A100 GPUs on GCP.
- **Clause Identification and Risk Analysis:**
- Extracted clauses with more than 85% accuracy, **by fine-tuning BERT models** for multi-label sentence classification.
- Performed risk analysis by comparing extracted clauses for compliance with preset clauses, with over 80% accuracy. For this we fine-tuned BERT model using **sentence-transformer architecture** to minimize the cosine similarity loss.
- Enhanced ML development efficiency and model reliability through **MLflow's** comprehensive experiment tracking, hyperparameter optimization, and artifact management capabilities for all BERT fine-tuning efforts.

Data Scientist Intern - Airtel - Gurgaon **May 2020 - June 2020**

- Performed **multivariate time series forecasting** to predict 4G data usage volume for network planning.
- Trained **decision-tree and XGB models** utilizing network usage statistics for anomaly detection in mobile towers.

Full Stack Developer (Freelance) - Go Jain Yatra - Mumbai **May 2018 – July 2019**

- **Designed, developed and deployed** a website for booking Jain Daharamshalas using Django, Bootstrap and PostgreSQL

Education

Master of Technology in Computer Science, *Indian Institute of Science, Bangalore* Aug 2019 - June 2021
Bachelor of Technology in Engineering Physics, *Indian Institute of Technology Delhi* Aug 2014 - June 2018

Skills

Languages, DBs & Cloud	Python, JavaScript, Typescript, C++, SQL, MongoDB, Qdrant, Pinecone, AWS, GCP, Azure.
Libraries & Tools	FastAPI, Flask, Langgraph, Tensorflow, PyTorch, MLFlow, Docker, React, Tailwind, Shadcn, Git

Academic and Personal Projects

Agentic RAG using LangGraph **May 2024 - July 2024**

- Implemented multi-agent Supervisor-Assistant architecture using LangGraph and UI using Chainlit. Created SQL agent to answer queries over multiple tables. Created Chart agent for plotting graphs and Report agent for generating reports.

Automate short form content creation with AI **Sept 2023 - Dec 2023**

- Used AI technologies such as LLMs to write a script, Stable Diffusion models for image generation, Text to Speech models for Audio generation, voice cloning and Whisper model for subtitle generation, to automate short video creation.

Graph Convolutional Network (GCN) for Text Classification **Oct 2020 – Jan 2021**

- Implemented text classification as node classification and graph classification tasks using GCN model. Achieved state of art classification accuracy of 96.8% and 92% on R8 and R52 datasets of Reuters, and 76.6% on Movie Review dataset.

Awards and Achievements

GATE exam: Secured an All India percentile score of 99.78 in Computer Science GATE exam, 2019. 2019
Scholarship for Higher Education (SHE), Stood within top 1% of School Board at class XII, 2014 2014